用Excel解決農業研究資料統計分析的方法(四)－相關與迴歸

生物統計

農試所農藝組 呂秀英 魏夢麗 呂椿棠

一、前　言

本系列報告在介紹如何使用Excel進行農業科學研究中資料分析的方法，前三篇已介紹了敘述統計量、t檢定、F檢定、χ^2檢定，而本篇重點則在說明如何利用Excel來進行相關與迴歸。

二、相關與迴歸的意義

利用兩個變數資料通常可以研究相關和預測的問題。相關是指兩個變數之間關聯的強度，瞭解相關通常有二種方式，一為繪製資料散佈圖，另為計算相關係數(correlation coefficient)。而當兩變數有相關時，則可進行簡單迴歸分析(simple regression analysis)，通常可由一個自變數(independent variable, 預測變數, X)，來預測一個依變數(dependent variable,被預測變數, Y)。當欲利用多個自變數(X_1,X_2,...)來預測一個依變數(Y)時，則使用複迴歸分析(multiple regression analysis)。

相關與迴歸是使用最頻繁的統計方法，但兩者間的意義常常被混淆不清。兩者之試驗目的不同：相關分析，全稱為簡單相關分析，是用來測定兩變數(無依變數與自變數之分)之間是否存在簡單的直線關係，當相關分析結果顯著，表示兩變數間存在直線關係，但若分析結果不顯著，就不能認定兩者間無任何關係或互相獨立，因為也有可能是其他二次以上的關係，故其正確結論是兩變數間無直線關係；迴歸分析則是用來描述依變數(Y)受到自變數(X)的影響程度，而希望能夠尋得一個方程式，使其能藉由X來預測Y。再者在統計上，相關與迴歸的資料性質也不同：用以進行相關分析的兩個變數都必須為不可自由變動的隨機變數，例如兩種調查性狀間之關係；而迴歸分析的Y為隨機變數但X卻最好是可自由變動或控制的固定值，例如乾物重受施肥用量之影響，乾物重為隨機變數而施肥用量為試驗者自行給定的固定值，因為這是在以最小平方法(least squared method)估計迴歸係數的傳統方法所要求的前提。若迴歸分析中的X為隨機變數，則迴歸係數之估計應改採用最大概度法(maximum likelihood method)，但一般統計軟體的內設都是使用傳統方法，除非去更改設定或另外購買進階模組，才能在最大概度法的模式下操作迴歸分析。

無論是相關或迴歸，解釋的是變異的問題，因此最好都不要將變異和因果混為一談。我們不能光憑相關或迴歸顯著存在，就認定X是造成Y結果的原因，因為兩者之間有可能是其他變數所引起的間接關係或影響。

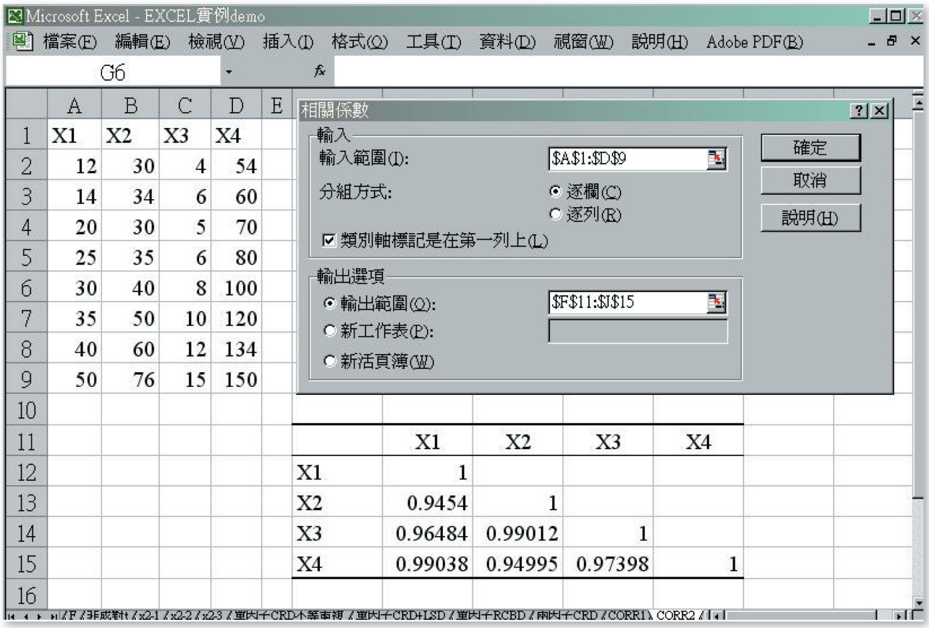
作者：呂研究員秀英
連絡電話：04-23302301-125

三、相關分析之求算步驟及結果解讀

Excel的「工具」中的「資料分析」提供了很多現成的統計分析工具，只需指出資料所在的單格和提供必要的參數，該工具就會使用適宜的統計或工程函數，對資料做處理，給出相應的結果，比起輸入公式和插入函數，更為簡便。但在安裝Excel 時，通常不會自動安裝此工具。若在「工具」中沒有發現「資料分析」，首先要勾選「工具」中的增益集，從其「增益功能」內勾選「分析工具箱」後便可完成安裝。Excel在「資料分析」裡提供了「相關係數」的計算，但不提供相關係數的顯著性測驗，不過對此我們仍然可以透過輸入公式或函數獲得。以4個變數間的相關分析為例說明，資料如圖一所示，首先進入「工具」中的「資料分析」選擇「相關係數」，其操作步驟如下(圖一)：

1. 指定「輸入範圍」－可用輸入範圍位址或用滑鼠拉選範圍；如本例指定的輸入範圍若包含變數標記在內，可直接輸入範圍為相對位址A1:D9。(倘使用滑鼠拉選範圍的方式，輸入範圍自動顯示絕對位址$A$1: $D$9)。
2. 依資料輸入格式選擇「分組方式為逐欄或逐行」－因本例為4個變數分欄輸入，故勾選逐欄。
3. 勾選「類別軸標記是在第一列上」－因本例在指定範圍時，將變數標記一起納入，故必須勾選此項；但若只有指定資料本身，未將標記納入，則此項毋須勾選。
4. 在「輸出選項」中勾選分析結果所在位置－可以自行決定要放在同一工作表內某處「輸出範圍」內、「新工作表」或「新活頁簿」。如本例希望將輸出結果和資料放在一起，故勾選輸出範圍並指定位址。
5. 最後按「確定」，即給出輸出結果－Excel會以矩陣形式來列出相關係數的結果。如圖一所示之本例結果。

由於Excel的「工具」中的「資料分析」並不提供相關係數的顯著性測驗，因



	A	B	C	D
1	X1	X2	X3	X4
2	12	30	4	54
3	14	34	6	60
4	20	30	5	70
5	25	35	6	80
6	30	40	8	100
7	35	50	10	120
8	40	60	12	134
9	50	76	15	150

	X1	X2	X3	X4
X1	1			
X2	0.9454	1		
X3	0.96484	0.99012	1	
X4	0.99038	0.94995	0.97398	1

圖一、用Excel「工具」中的「資料分析」求4個變數間的相關係數。

此必須接著以輸入公式或函數來進行。相關係數(r)的顯著性測驗可利用t檢定，公式如下：

$$|t| = \frac{|r|}{\sqrt{\frac{1-r^2}{n-2}}} > t_{(\alpha/2,n-2)}$$

式中n為樣本數，$t_{(\alpha/2,\ n-2)}$為α顯著水準下自由度等於n-2的雙尾 t 臨界值。t 統計值的計算，可利用Excel的TINV函數，其格式為：TINV(probability, degree_freedom)，probability為雙尾機率值(故若指定顯著水準為0.05，只需直接輸入0.05即可，不必再除以2)，而degree_freedom為自由度。而Excel的絕對值及開方根函數分別為ABS及SQRT。因此，以本例的X1和X2兩變數間的相關係數0.9454而言（位址在G4)，自由度=8-2=6，其t檢定的實測值和臨界值的Excel公式寫法如下：

t 實測值= ABS(G4)/(SQRT(1-G4*G4))/6)

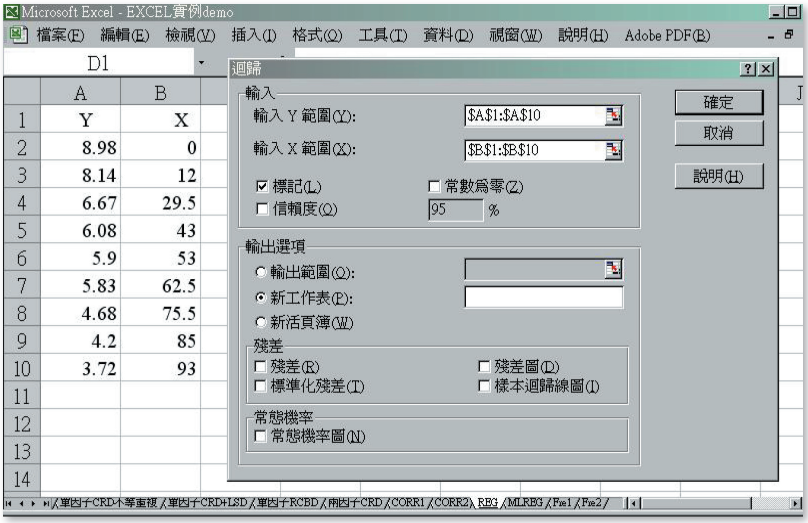
t 臨界值= TINV(0.05, 6)

各得到7.1055和2.4469，t 的實測值大於其臨界值，表示X1和X2兩變數間存在顯著的正相關關係。

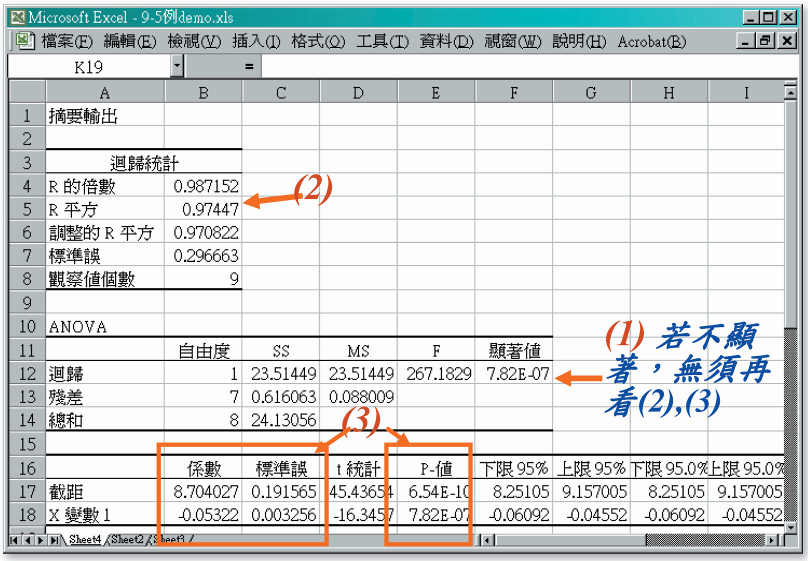
四、簡單迴歸分析之求算步驟及結果解讀

假設欲探討相對濕度(X,%)增加對某揮發性化合物之量損失(Y,mg)的影響程度，其資料輸入如圖二所示。進入Excel的「工具」中的「資料分析」選擇「迴歸」，其操作步驟如下(圖二-(A))：

(A)



(B)



圖二、用Excel「工具」中的「資料分析」進行迴歸分析。(A)操作畫面，及(B)輸出結果。

1. 指定「輸入Y範圍」和「輸入X範圍」－可用輸入範圍位址或用滑鼠拉選範圍；如本例指定的輸入範圍若包含處理名稱在內，可直接輸入範圍的相對位址，如A1:A10和B1:B10 (倘使用滑鼠拉選範圍的方式，輸入範圍自動顯示絕對位址$A$1: $A$10和$B$1: $B$10)。
2. 勾選「標記」－因本例在指定範圍時，將變數名稱(標記)一起納入，故必須勾選此項，否則毋須勾選。
3. 在「輸出選項」中勾選分析結果所在位置－可以自行決定要放在同一工作表內某處「輸出範圍」內、「新工作表」或「新活頁簿」。本例希望將輸出結果置於新的工作表內，故勾選新工作表。
4. 至於「常數為0」、「信賴度」、「殘差」、「常態機率」等為自由選項－但注意Excel所求得的標準化殘差之結果不可信，故建議勿選該項。
5. 最後按「確定」，即給出輸出結果－若不指定自由選項，Excel會列出迴歸統計值(內含決定係數R^2)、變方分析(ANOVA)、截距和斜率估計值及其上下限和顯著性測驗；若還指定了其他自由選項，則會進而列出預測值及其殘差、迴歸圖。如圖二-(B)所示，判讀迴歸分析的輸出結果，有三個步驟：(1)首先檢視變方分析結果是否顯著，若不顯著，表示X和Y間的迴歸關係根本不存在，因此無須再繼續觀看其他輸出結果；(2)在變方分析顯著之情形下，檢視R^2值大小，該值愈大，表示用來判斷迴歸估計式的解釋能力愈好，即對該模式愈滿意；(3)若R^2值夠高，再檢視斜率(迴歸係數)是否顯著異於0，若顯著表示Y會隨著X而呈直線遞減或遞增；若不顯著則意謂Y不因X而異，即該迴歸式不具意義。本例結果顯示，X與Y間的迴歸關係極顯著存在，且模式的解釋能力也甚高，迴歸估計式可寫成$Y=8.7040-0.0532X$，$R^2=0.97$。該迴歸式的意義解讀為表示每增加一單位相對濕度，將約損失0.0532單位的重量。

五、複迴歸分析之求算步驟及結果解讀

Excel的「資料分析」的「迴歸」，也可用來進行複迴歸分析，其操作方法完全一樣，只要在「輸入X範圍」內納入欲分析的多個自變數即可，但這些自變數欄位間必須是連續的。複迴歸的輸出結果，除了自由選項的標準化殘差不可信以外，其迴歸圖也是錯誤的，千萬別採用。判讀複迴歸分析的輸出結果，一樣循著上述三個步驟，即能將顯著存在的複迴歸估計式整理出來。

六、二次曲線迴歸分析之求算步驟及結果解讀

只要稍加運用一些技巧，Excel的「資料分析」的「迴歸」一樣可以用來進行二次曲線迴歸分析。二次曲線迴歸與複迴歸之差異在於，前者只有一個自變數X，但與依變數Y呈二次曲線(拋物線)關係，而後者是多個自變數對Y之影響探討。由於二次曲線迴歸式為$Y=b_0+b_{11}X+b_{12}X^2$，若將該式的二次項變數X^2視為第二個自變數，這個新變數的數值由原變數數值取平方而得，如此在利用Excel進行分析時，就可以完全仿照複迴歸分析的操作方法，於「輸入X範圍」內指定兩個自變數欄位，即可獲得迴歸分析的結果；在輸出結果中第二個變數的迴歸係數就是二次項變數X^2的迴歸係數。同理，當然Excel的「資料分析」的「迴歸」還可運用到其他可經適當變數換後能改寫成直線形式的曲線迴歸模式。